The Effectiveness of Social Norms in Fighting Fake News on Social Media

Henner Gimpel, Sebastian Heger, Christian Olenberger & Lena Utz

Routledge
Taylor & Francis Group

Check for updates

The Effectiveness of Social Norms in Fighting Fake News on Social Media

Henner Gimpel[a], Sebastian Heger[b], Christian Olenberger[b], and Lena Utz[b]

[a]University of Hohenheim, FIM Research Center, Project Group Business & Information Systems Engineering of the Fraunhofer FIT, Stuttgart, Germany; [b]FIM Research Center, University of Augsburg, Project Group Business & Information Systems Engineering of the Fraunhofer FIT, Augsburg, Germany

ABSTRACT

Fake news poses a substantial threat to society, with serious negative consequences. Therefore, we investigate how people can be encouraged to report fake news and support social media platform providers in their actions against misinformation. Based on social psychology, we hypothesize that social norms encourage social media users to report fake news. In two experiments, we present participants a news feed which contains multiple real and fake news stories while at the same time exposing them to injunctive and descriptive social norm messages. Injunctive social norms describe what behavior most people approve or disapprove. Descriptive social norms refer to what other people do in certain situations. The results reveal, among other things, that highlighting the socially desired behavior of reporting fake news using an injunctive social norm leads to higher reporting rates for fake news. In contrast, descriptive social norms do not have such an effect. Additionally, we observe that the combined application of injunctive and descriptive social norms results in the most substantial reporting behavior improvement. Thus, social norms are a promising socio-technical remedy against fake news.

Introduction

"Once upon a time … " is a standard opening of a fairy tale. How easy it would be if there was such a striking characteristic marking fake news! Fake news "are intentionally and verifiably false" [3, p. 4], like a fairy tale. Fairy tales are poignant stories that spread widely and last a long time. In a similar vein, fake news spread "significantly farther, faster, deeper, and more broadly" than real news [87, p. 1147]. However, unlike a fairy tale, fake news does not readily disclose its inauthentic nature.

Fake news on social media is a severe problem for society. More and more people search and consume news stories from social media instead of traditional news media [35, 61, 79]. Social media are especially vulnerable to fake news, as information verification—traditionally part of journalistic efforts—cannot be ensured [72]. There are many examples of fake news in domains such as public health (e.g., incorrect information known as infodemics

during the COVID-19 pandemic [48]), climate change [85], politics [2], or stock markets [18]). Fake news stories lead individuals to believe inaccurate information and influence their opinions and actions [3].

Social media platform providers are under increasing pressure to take more proactive actions to stop the spread of fake news. Many social media platforms provide their users with the option to report fake news so that the fake news can be professionally checked, flagged, or even removed. However, anecdotal evidence suggests that users only sporadically report news stories as fake. Findings of our empirical research support that users never or only rarely report fake news. Thus, while a large user base generates the content on social media, and the social media platforms allow their users to contribute to keeping social media free of fake news, the users hardly take up this possibility. "Clearly, helping users make a more informed decision on social media is an important piece of the solution in curbing the spread of fake news" [44, pp. 1025-1026] and supporting platform providers to comply with their responsibility. The question arises, how to motivate users to report fake news?

Social psychology suggests that social norms (SNs) strongly influence behavior. They might be a solution to the riddle. SNs can steer behavior in a helpful direction for individuals and society [13, 15, 17]. There are two types of SNs: Injunctive SNs describe which behavior most people approve or disapprove. Descriptive SNs refer to what other individuals do in particular situations [17]. Previous research showed that SNs are a powerful tool to encourage prosocial behavior in various domains [13, 17, 34].

However, it is not clear whether social media is an adequate means to convey SN messages. First, the bystander-effect suggests that individuals react reservedly when other individuals are present because of a missing sense of responsibility [50], also in the context of social media [25]. Second, SN messages can backfire [13]. On the one hand, social media providers may apply untargeted descriptive SNs, which indicate that most users do not report fake news and, thus, demotivate users. On the other hand, SN messages may decrease news articles' believability, leading to increased reporting of real news [76]. Hence, it is a mystery whether and how far SN messages apply to the context of fake news reporting in social media and, if they do, whether their net effect is positive or negative.

Information systems (IS) scholars recently focused on multiple cognitive and behavioral aspects of users consuming fake news on social media. This stream of IS research suggests that subtle changes to the user interface may trigger reflection and lessen unwanted user behavior contributing to the further spread of fake news [44, 45, 54–56]. These results are encouraging for user interface design elements such as SN messages to steer user behavior. However, none of the recent IS papers investigated social norms or other means to improve social media users' reporting of fake news. Against this background, we pose the following research question:

Can injunctive and descriptive social norm messages provided as part of a social media user interface improve social media users' fake news reporting behavior?

To answer our research question, we derive theoretical hypotheses and conduct two online experiments. Based on a combination of theoretical arguments and empirical evidence, we suggest that combining injunctive and descriptive SN messages as social

media design features has positive effects on fake news reporting. Interestingly, descriptive SNs alone do not affect the reporting of fake news.

Our results contribute to research and practice. From a scholarly perspective, we follow a recent call for more IS research on socio-technical interventions as remedies against fake news [8] and contribute to the emerging stream of IS research on fake news [e.g., 44, 45, 56]. From a practical perspective, we provide evidence for SNs as a potential tool for social media platform providers in their fight against fake news. Ultimately, using this tool may contribute to less fake news and, thus, lessen the societal burden of social media's adverse effect.

Theoretical Background and Hypotheses Development

Fake News

Fake news in itself is not a new phenomenon. In the past, various groups of people, such as advertisers, political activists, and religious zealots, have used different forms of fake news to influence public opinion or spread propaganda [59, 79]. Two key characteristics make news stories fake news: inauthenticity and deception. On the one hand, fake news contains verifiably false information that is nevertheless deliberately disseminated [79]. On the other hand, fake news can mislead consumers, whereby incorrect information is interpreted as truthful [3, 59]. Fake news appears in the form of text, images, audio files, or videos that are edited or deliberately taken out of context [59, 79].

Besides the different forms of fake news, there are also different motivations for individuals and organized groups to produce and distribute false information. First, fake news, such as satirical content or hoaxes, serves as a source of entertainment. Second, due to their eye-catching and sensation-oriented design, fake news is well suited to induce readers to visit the original source [72, 79] and thereby generate a financial profit through advertising [3]. Third, fake news can influence opinion. The aim may be to create an advantage for the favored political candidate [3, 4, 59, 79, 88] or to influence public opinion on a political or social issue. Therefore, fake news's main danger lies in the fact that people do not recognize the false information and the associated manipulation. This misinformation could lead to an unjustified change of voting behavior or political or social attitudes resulting in a severe negative impact on individuals and society.

Even correcting misinformation does not necessarily change people's beliefs [31, 63]. Furthermore, even when an individual believes the correction, the misinformation may persist. Any misinformation repetition can have negative consequences, even in the context of refuting it [36]. This effect is due to familiarity and fluency biases. The more people hear a story, the more familiar it becomes, and the more likely they are to believe it as real [37, 67, 77]. As a result, exposure to misinformation can have long-term effects, while corrections may be short-lived. Hence, limiting the spread and impact of fake news is essential.

Fake News in Social Media

Although fake news has already been used in the past, its significance and dissemination have recently gained importance [78]. Social media, in particular, provides a platform for the effective spread of fake news. First, social media offers a broad audience [26, 84]. For

substantial parts of this audience, social media is an essential source of news [35, 61], which makes its users highly vulnerable to fake news. Second, the barriers to enter social media are low [3]. For instance, the low effort involved in setting up an account also enables the creation of machine-controlled user accounts, such as social bots [28]. As a result, about 19 million machine-controlled accounts were active on Twitter during the U.S. presidential election 2016, spreading information about the candidates [64]. Third, transmission speed and reach through social media are higher than ever before. Since exchanging information is an essential part of social media [43], it is easier to discuss or share both real and fake news stories [79]. About 25 percent of U.S. adults reported that they shared fake news deliberately or unknowingly [6]. Therefore, information can quickly reach global access virally [78, 90]. The novelty of fake news and the negative feelings they often cause are additional reasons that fake news travels faster than real news [87]. Simultaneously, the decentralized distribution makes it difficult to stop the dissemination of fake news since there is no longer a central origin of spread that can be addressed. Fourth, no authority controls the authenticity of information [78]. The process of journalistic information verification and the associated gatekeeping functions can no longer be guaranteed because social media allows everyone to participate in news production [2, 65]. Due to the lack of control, validating the authenticity of information now lies with the users [72]. Furthermore, the widespread dissemination of fake news may reduce the demand for authentic information [3].

However, people's ability to decide whether the information is true or false has some natural weaknesses, especially in the social media environment [51, 56, 79]. A key factor for validating authenticity is the credibility of information, which is complex and diverse [68, 89]. Different factors, such as source (expertise, knowledge, and trustworthiness), receiver (receptivity for a message), and message characteristics (consistency and clarity) all interact in the assessment of the credibility of information [89]. Additionally, the beliefs and prior knowledge [67] of a user are important because these factors can lead to confirmation bias. People prefer to believe information that matches their preexisting beliefs [3, 38, 44]. When they encounter information that challenges their preexisting beliefs, they experience cognitive dissonance [29, 53]. Resolving this cognitive dissonance takes cognitive effort, and people tend to avoid this [82]. Simultaneously, the assessment of other social media participants also impacts the credibility of news articles [45]. Due to these factors, it is necessary to develop countermeasures that support social media users in handling fake news.

From a different theoretical perspective, previous research on the bystander effect shows that people's willingness to help is reduced the more other people are present [30]. This effect applies not only in the case of physical presence [50, 74] but also in an online environment [30]. Following the bystander intervention model, a bystander must follow five steps: noticing and appraising an event as an emergency, recognizing the own responsibility, deciding to intervene, and finally intervening [43]. These steps can be affected by the presence of other people. For instance, a larger group size leads to reduced response rates and willingness to help in online communities [86] or during e-mail communication [5, 11]. In social media, where many users see the same content, this could lead to a diffusion of responsibility [86] where the individual does not take responsibility for reporting fake news. The user not feeling accountable is another factor contributing to this effect. Within a group of people, an individual evaluates its accountability lower because another individual may be more accountable to intervene in a specific situation. Making accountability transparent

may reduce this effect. However, transparency can also negatively affect when users feel that their intervention is evaluated by others and, therefore, do not dare to take action [25]. For these reasons, and because of the importance of the other countermeasures, it is necessary to show social media users that reporting is desirable.

Countermeasures Against Fake News

A body of IS research is emerging that addresses questions around the consumption of fake news [8]. Kim and Dennis offered evidence that changing how the source of news is presented leads social media users to better reflect on whether to believe articles and perform activities that contribute to their spread [44]. Specifically, both changing the interface to highlight the article's source and providing source ratings by a board of experts positively affected users' cognitive and behavioral processes. In a similar vein, Kim et al. investigated three different types of source reputation ratings and suggested that these ratings influence the extent to which users believe articles on social media [45]. Moravec et al. found evidence that the mere fact of asking people to reflect on their background knowledge and rate the truthfulness of a news story made them think more critically about that news story and others [54].

Despite these results, in practice, many social media platform providers are criticized for the lack of countermeasures against fake news [83]. Therefore, social media providers tried to develop various countermeasures to prevent the spread of fake news, such as automatically detecting social bots or cyborg users. Another method is to reduce the financial incentives to spread fake news by prohibiting sites from advertising if they have repeatedly shared fake news [80]. However, current approaches do not automatically check the authenticity of information [90]. Due to fake news' complex structures, fake news detection is challenging to automate [12, 51, 79]. Therefore, social media providers work with third-party fact-checking organizations who manually check suspicious information [40]. The social media provider then can flag the corresponding content and display other articles on this topic.

However, in a study on social media interface design, Moravec et al. investigated the effects of flagging fake news as "disputed by 3rd party fact-checkers" [56, p. 1345]. They found that flagging fake news triggered increased cognitive activity but did not affect judgments about truth and users' beliefs. In their experiment, the fake news flag was not strong enough to make users overcome their a priori beliefs and confirmation bias. In a recent study, Moravec et al. investigated effective approaches to implementing fake news flags by adapting them to cognitive processes [55]. Their results indicate that the specific design of fake news flags has a significant influence on the believability of the news article [55].

All of these approaches depend on a pre-selection of suspicious content. Therefore, social media providers rely on the support of their users [58]. For example, Facebook sends content to third-party fact-checking organizations for verification if enough users report specific content as fake [57]. The users and their reporting behavior become a central element of the system because they act as enablers for the subsequent mechanisms. However, no data on users' reporting behavior is available or publicly accessible [59].

Social Norms

A notable characteristic of social media is that decision-making is not isolated but occurs in a virtual social environment. Within such a social setting, others' attitudes and behaviors can result in SNs that strongly influence an individual's behavior [17, 24]. SNs are "an individual's beliefs about the typical and condoned behavior in a given situation" [46, p. 480]. SNs are an essential factor in the decision-making process and can lead to prosocial human behavior [13, 17, 34]. They may be communicated via SN messages visualizing or verbalizing the norm. Research distinguishes two types of SNs. First, injunctive SNs "specify what ought to be done" [17, p. 1015] and describe the behavior, which is expected from the social environment by defining which actions are desired and which are undesired [69]. Second, descriptive SNs inform about what others typically do in a particular situation [17]. Although both SNs provide social information, they act through different motivations [17, 69, 70]. Injunctive SNs influence behavior by highlighting potential social rewards and punishments for desirable or undesirable behavior [16]. Such moral evaluation influences the behavior even if the imaginary others are not family or friends but are general members of society [14].

Although SNs mostly guide an individual's behavior, they do not always, and in every situation unfold the desired effect [17]. The focus theory of normative conduct [17] assumes that an individual is more likely to behave according to a SN when the individual's attention is focused on the norm at the moment of behavior. Priming or highlighting other individuals' behavior are exemplary tools to activate a SN [17, 34].

Previous research already demonstrated the importance of injunctive SNs for decision-making in different contexts such as alcohol use [49, 70], gambling [60], tax evasion [92], organ donation [65], and risky sexual online behavior [7]. Due to these versatile influences on behavior, we suspect that injunctive SNs can also be used to improve social media users' reporting behavior. We expect that when social media users are made aware that reporting fake news is a behavior desired by the social environment, it also positively affects their decision to report fake news. In particular, we expect that the use of an injunctive SN will lead to more reported fake news stories. Based on this, we derive the following hypothesis:

Hypothesis 1 (H1): The presence of an injunctive SN message indicating that reporting fake news is a socially desirable behavior increases the amount of reported fake news.

Contrary to injunctive SNs, descriptive SNs motivate actions by providing information about others' behavior and showing which actions are likely to be effective, adaptive, and appropriate [16]. Others' behavior can be used as a heuristic, giving the decision-maker an information-processing advantage and decisional shortcut in uncertain, ambiguous, or threatening situations [42]. Previous research has demonstrated the successful application of descriptive SNs to reduce littering [17, 69], support recycling [62], improve voting behavior [32], promote towel reuse [34], increase monetary donations [14, 20], and to encourage healthier eating [71]. Based on these results, we suspect that the demonstration of others reporting content on social media platforms improves users' reporting behavior. Therefore, we suspect that the number of reports is increasing compared to no SNs message. This assumption leads to the following hypothesis:

Hypothesis 2 (H2): The presence of a descriptive SN message indicating that other people report fake news on social media increases the amount of reported fake news.

In addition to the isolated consideration of the two types of SNs, a combined application is also possible. In a study on recycling behavior, Cialdini [13] demonstrates that the joint use of injunctive and descriptive SNs leads to a particularly high intention to recycle. Similarly, Schultz et al. [75] reveal in a field experiment on towel reuse that combining the two types of SNs has a more substantial impact on behavior than just using one of them. Thus, previous research shows that although both types of norms use different motivations [17, 69, 70], their simultaneous application can produce additional positive behavior effects. Focus theory suggests that the presence of a descriptive norm message can additionally direct the user's focus to the injunctive norm [17]. Applying these insights to the case of fake news reporting, we expect that the presence of a descriptive SN message—describing other users have reported a news post—will further focus a user's attention to the injunctive SN—describing that reporting is a desired behavior—which consequently increases the number of fake news stories reported. Therefore, we expect that the number of reports will exceed the number of reports when only applying one SN message at a time (i.e., injunctive or descriptive). Hence, we formulate the following hypothesis:

Hypothesis 3 (H3): The simultaneous presence of injunctive and descriptive SN messages indicating that reporting fake news is a socially desirable behavior and that other people are reporting fake news leads to a higher amount of reported fake news than the use of only one of the two types of SNs.

However, a particular characteristic of descriptive SNs is that a descriptive SN's strength can be varied by adjusting the number of people who show a specific behavior. Current research demonstrates a positive correlation between the strength of a descriptive SN and the influence on behavior [21, 46]. For example, Demarque et al. [21] reveal in an online shopping environment that information about a higher number of people who bought an ecological product leads to more other customers buying ecological products. The higher the number of people who demonstrate a specific behavior, the more this behavior is perceived as the standard way of doing things and increases the willingness to use the SN as a decisional shortcut. In combination with an injunctive SN, increasing the strength of a descriptive SN further focuses the attention of people on a desirable behavior (represented by the injunctive SN) and motivates them to behave in the desired way [76]. Vice versa, a weak descriptive SN indicating that most people do not act desirably may be counter-productive [17].

For fake news, we suspect that the presence of a stronger descriptive SN is focusing a user's attention on a specific post and the action of reporting. Simultaneously, the descriptive SN highlights the injunctive SN and increases the likelihood that a user is reporting a specific fake news post. Therefore, we expect that the more users have reported a particular fake news story, the greater is the likelihood that other users will also report this fake news story. However, the results of Wong et al. [94] show that the intention to report depends on the perceived benefit of the reporting action. Users report more often when they feel that reporting also helps to solve the problem [94]. Therefore, we suspect that an increase in the strength of the descriptive norm could lead to the user perceiving a little

additional benefit in his or her reporting action since enough others have already reported a specific fake news story to draw the social media's attention provider. This effect counteracts the positive influence of both the descriptive and injunctive SN. We suspect that from a specific strength onward, the reduction due to the low perceived additional benefit surpasses the increase due to the descriptive SN. If this threshold is exceeded, a further increase in strength leads to a reduction in reporting probability. Based on this, we deduce the following hypothesis:

Hypothesis 4 (H4): The strength of a descriptive SN message—while an injunctive SN is also present—affects a user's likelihood to report a news story as fake. The influence is non-monotonic and follows an inverted u-shape.

Despite the desired effect of SNs (that fake news is reported), it should be noted that also undesired effects of SNs may occur. Specifically, users might report truthful news as fake. We do not hypothesize any effects here, as this is not the focus of the present article. However, we include this undesired effect on reporting behavior in our analyses to test whether the hypothesized benefit of SNs comes at a price in terms of misleading reports.

Study 1

To test our theoretical hypotheses, we conducted two online experiments, Study 1 and 2. The studies are closely related, with Study 2 delving deeper on H2.

Method

Participants

We recruited participants for our experiment via Facebook groups for students. Students are an exciting group as they use social media regularly, in many cases daily [66], which makes them particularly open to the frequent exposure to fake news. Using an online (rather than a lab) experiment and sampling via Facebook are sometimes discussed as methodological weaknesses. We instead consider these design choices as strengths supporting the experiment's ecological validity for our research context. We recruited participants via German Facebook posts. The experiment was implemented in German. Screenshots presented here are a translation. We motivated participation by a raffle for three vouchers.

To ensure the responses' quality, we include only data from participants who use social media at least once a week and who completed the experiment. Explicit checks were made to ensure that participants interacted with all the experiment materials to see all the content. This approach leads us to data from 320 active social media users. The sample is strongly composed of students (n = 293) with an average age of 23.2 years. A total of 70 percent of the participants are female (n = 224). In total, 267 participants state that they use social media as a source of information on current events, public affairs, and political issues at least once a week. Previous studies also highlight the importance of social media as a news source [61]. The participants rarely report posts (mean 0.12 per day, a standard deviation of 0.51). The Online Supplemental Appendix A has details.

Task

We provide the participants with a self-developed news feed to manipulate and monitor the experimental environment. To increase the degree of realism and create a familiar environment for participants, we orient our development towards the best-known news feed design developed by Facebook [61]. We display content in the form of posts containing a source, an image, a headline, and a lead sentence. We exclude other information that could be used by the participants as potential references to evaluate the truthfulness of a news story according to its general acceptability (social influence bias), such as the number of likes, comments, and shares. Also, every post has three standard Facebook buttons for interaction: a like button, a comment button, and a share button. We have added a report button. Such a reporting feature already exists on Facebook, but it requires several clicks to use it. We have simplified the reporting process to reduce possible difficulties and, therefore, to separate the different treatments' effects better. Figure 1 shows an exemplary implementation of a post in the experiment.

The news feed consists of 15 posts of three different types. First, the news feed includes five fake news posts containing fake news stories that spread widely on Facebook (in German-speaking countries) and are proved to be false [74, 81]. We modified the fake news posts based on common fake news characteristics such as spelling mistakes, flashy formatting, or obviously altered images to make it easier to identify them. Every fake news story is assigned to a non-existent news source to reduce authenticity further. Besides the fake news, the news feed includes five real news posts. These contain no characteristics of fake news and originate from authentic news sources. In addition to the news posts, the news feed also has five neutral posts of imaginary friends to create an environment that is somewhat realistic and increases external validity. The display order of the posts is generated randomly for each participant. The Online Supplemental Appendix B contains all posts.
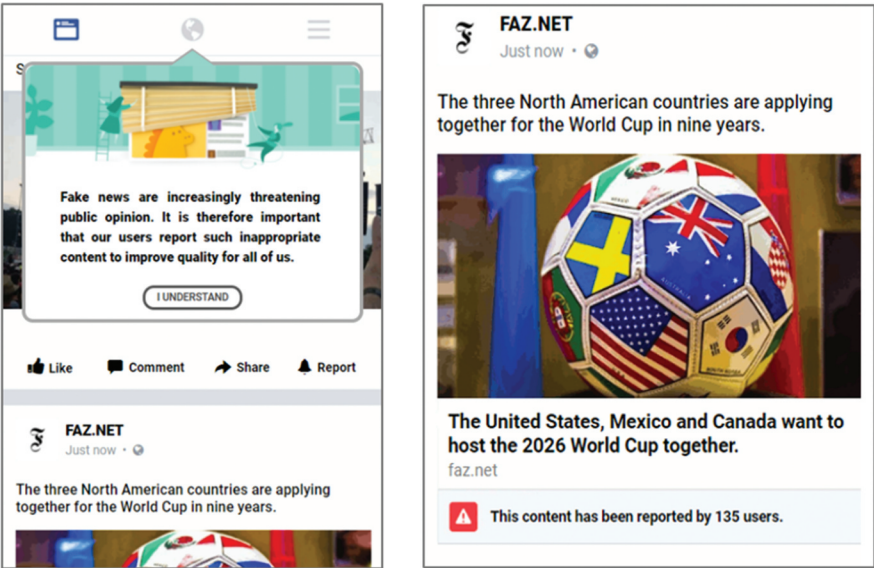

Figure 1. Implementation of the injunctive social norms (SN) message (left) and a descriptive SN message (right).

Treatments

We confront participants with different SNs and observe fake news reporting behavior. We use a 2 (injunctive SN message: present vs. absent) x 2 (descriptive SN message: present vs. absent) between-subject design:

(1) *Control treatment*: No SN messages were used in the news feed.
(2) *Injunctive SN treatment*: The injunctive SN message was shown at the beginning of the news feed.
(3) *Descriptive SN treatment*: Five of the ten posts with news content, whether they are fake news or real news, were randomly flagged with a descriptive SN message.
(4) *Combined treatment*: Combination of (2) and (3).

To vary the descriptive SN's strength, we use a 5-level within-subject design in the treatments with descriptive SN messages. Each level occurred exactly once per participant.

Based on random allocation, the number of participants per treatment ranged from 77 to 83. To test the structural equality of the four treatments, we conduct ANOVA tests (for age and usage behavior of social media) and chi-squared tests (for gender, education, and occupation) [52]. The results indicate no differences at a 5 percent significance level. Thus, we assume structural equality. A detailed list of the results can be found in Online Supplemental Appendix C.

Independent Variables

To be effective, SNs should be perceived as appropriate and credible [20], and they should be present in perception or the focus of attention [16]. Based on these recommendations, we developed implementations for injunctive and descriptive SN messages. We paid particular attention that the norm messages do not appear like an instruction by the experimenter but as part of the social media news feed. The injunctive SN message is displayed as a text at the beginning of the news feed (left-hand side of Figure 1). The text reads as follows: "Fake news is increasingly threatening public opinion. It is, therefore, important that our users report such inappropriate content to improve quality for all of us." The participants have to confirm the text by clicking on the corresponding button to see the news feed. Thus, we can ensure that participants recognize the injunctive SN (the focus of attention).

We implement the descriptive SN message as additional information for a post, indicating how many users have already reported this post (right-hand side of Figure 1). By increasing the number of reports, the strength of the descriptive SN also increases. Our experiment distinguishes five distinct levels (5, 25, 125, 625, 3,125) to cover a broad spectrum without reducing credibility. The assignment of the strength of the descriptive SN to a post is random. Additionally, the number of reports for each post randomly increases by up to a maximum of 10 percent to ensure that our results are not based on a specific number of reports and that the underlying pattern and prominence of numbers are hidden from the participants.

Prior research consistently suggests a decisive role of confirmation bias in assessing social media posts [44, 45, 54, 56], which is a prerequisite for reporting fake news. We assume that confirmation bias will also be present in our experiment. Given the broad range of topics covered by the posts in our experiment, identical posts in all treatments, and random allocation of participants to treatments, we assume about identical prior beliefs and

knowledge on the posts in all treatments. Thus, confirmation bias cannot be the source of treatment differences in participants' behavior.

Dependent Variables

As the primary dependent variable, we measured how much fake news a person reported. Additionally, we also measured how much real news a person reported. This approach gives us an indication of how well the fake news was recognizable as such. For statistical tests, we apply a 5 percent significance level.

Procedures

Participants were informed that the aim was to investigate the behavior of users in social media. We deliberately avoided mentioning fake news. Participants were asked how often they use social media as a news source and how much they trust this information. Then, the participants went through an interactive tutorial, during which they read a description, explicitly explaining the different features available in the news feed. The participants interacted with a single news post only used for this introduction and did not indicate a social norm. The tutorial aims to increase the participants' awareness of the reporting feature and teach them how to report content without telling them that they should report.

Next, in the so-called natural interaction run, the participants saw the news feed with all 15 posts in random order and were asked to interact with it as if they do with their personal news feed. It is only in this phase that the four treatments differ. The natural interaction run is of the highest interest for testing our hypotheses.

Next, in the fake assessment run, the participants saw the news feed again, but all interactions were reset. The participants received explicit instruction to report fake news in the news feed and were incentivized by a higher probability of winning a voucher when performing well, specifically in the fake assessment run. The fake assessment run should not be overrated for potential effects from the multiple exposure and order of posts and the changed incentive compared to the natural interaction run. We merely use it to assess whether participants could differentiate between fake and real news if they paid close attention. As expected, the participants identified both the fake news and the real news (see Online Supplemental Appendix D).

The experiment ended with brief questionnaires on social media usage and demographics and a debriefing. The debriefing informed the participants that we had manipulated different features of the posts and that these do not necessarily correspond to the real world. Figure 2 summarizes the process of the experiment. We performed a pre-test (Online Supplemental Appendix D) and report all scales in Online Supplemental Appendix E.

Results

To measure the effect of SN messages on fake news reporting, we compare the number of reported fake news within the four treatments employing regression analysis. Due to the discrete nature of the dependent variable (integers in the range from zero to five) and non-normally distributed residuals in linear regression, we apply an ordered logistic regression [1]. The assumptions of an ordinal dependent variable, at least one categorical
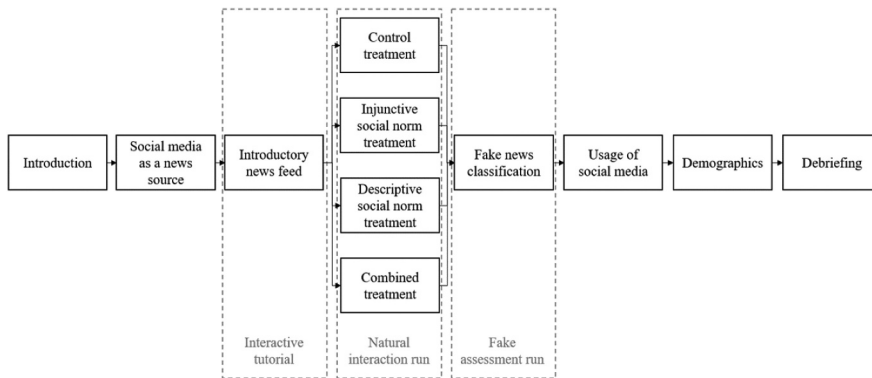
Figure 2. Summary of the procedure of the experiment (Study 1).

independent variable, and no multicollinearity are fulfilled by design. Additionally, we performed a Brant test and confirmed that the proportional odds assumption is satisfied (see Online Supplemental Appendix F).

We use the control treatment as the baseline to analyze the differences between the control treatment and the SN treatments. We perform a rebase to use the combined treatment as a baseline to investigate the differences between the combined treatment and the injunctive SN treatment and the descriptive SN treatment. Table 1 summarizes the results.

Table 1. Results of the ordered logistic regressions to compare the four treatments concerning fake news reporting in Study 1.[a]

Base Treatment			Coefficient	Standard Error	p-value		Odds Ratio (Effect Size)	Related Hypothesis
		Injunctive SN	0.674	0.326	0.039	*	1.962 (small)	H1
	Treatments	Descriptive SN	0.456	0.335	0.173		1.578 (small)	H2
		Combined	1.373	0.321	< 0.001	***	3.947 (medium)	H3
Control		0 | 1	0.927	0.249	< 0.001	***		
		1 | 2	1.793	0.263	< 0.001	***		
	Intercepts	2 | 3	2.312	0.276	< 0.001	***		
		3 | 4	3.406	0.328	< 0.001	***		
		4 | 5	4.558	0.451	< 0.001	***		
							Nagelkerke's R^2: 0.068	
		Control	-1.373	0.321	< 0.001	***	0.253 (medium)	
	Treatments	Injunctive SN	-0.699	0.292	0.017	*	0.497 (small)	H3
		Descriptive SN	-0.916	0.303	0.003	**	0.400 (medium)	
Combined		0 | 1	-0.446	0.209	0.033	*		
		1 | 2	0.420	0.209	0.044	*		
	Intercepts	2 | 3	0.940	0.217	< 0.001	***		
		3 | 4	2.032	0.270	< 0.001	***		
		4 | 5	3.185	0.409	< 0.001	***		
							Nagelkerke's R^2: 0.068	

Notes: [a]In the first model, the control treatment is the baseline, whereas, in the second model, the combined treatment is the baseline. $+ p < 0.1$; $* p < .05$; $** p < .01$; $*** p < 0.001$. n = 320.

H1: Effect of Injunctive SN Messages on Fake News Reporting

The regression analysis shows an injunctive SN message significantly increases the probability of fake news posts being reported. Compared to the control treatment, we observe that the odds of reporting fake news significantly increase by 96 percent (i.e., it almost doubles). Transforming the odds ratio to Cohen's d [73] and applying his standard levels for small, medium, and large effects [19] suggests that this effect is significant but small. As we hypothesized, pointing out the importance and the desirability in the social environment of reporting fake news guides the participants' behavior toward reporting more fake news in their news feed. The results support H1. An additional exploratory analysis of the data shows no significant effect on real news reporting while applying the injunctive SN message (see Online Supplemental Appendix G).

H2: Effect of Descriptive SN Messages on Fake News Reporting

Comparing the descriptive SN treatment to the control treatment, we see a 58 percent increase in odds of reporting fake news, but this small effect is not statistically significantly different from zero. H2 is not supported.

The data suggest a potential effect of the descriptive SN on the (erroneous) reporting of real news as supposedly fake. The odds of reporting real news increase by 232 percent. However, this increase also fails to reach statistical significance (p-value of 8 percent). Details on the regressions for real news reporting are in Online Supplemental Appendix G.

Not supporting H2 is counterintuitive. One reason could be that the descriptive SN was not sufficiently in the focus of attention as – contrary to the injunctive SN – participants did not have to acknowledge it. This idea is supported by the fact that only 25.1 percent of all participants who saw a descriptive SN message stated that they used the descriptive SN messages to identify fake news in the ex-post questionnaire. Another reason could be the different effects of positive and negative descriptive SNs [13, 15]. The news feed contains ten news posts, five real and five fake. In Study 1, the descriptive SN messages randomly flag five out of these ten news posts as having been reported by other users irrespective of whether the news is real or fake. Participants seeing fake news being flagged might perceive this as positive SN and might report fake news themselves. Participants seeing real news being flagged might perceive this as negative SN and might either also report real news themselves or might not engage in reporting at all. As the random allocation of the descriptive SN messages results in a mixed positive and negative SN message for most participants, we cannot single out the potentially contradictory effects of positive and negative descriptive SN with the data from Study 1. In Study 2, we separate the two types of descriptive SNs and investigate these potential effects.

H3: Effect of Combined Injunctive and Descriptive SN Messages on Fake News Reporting

Comparing the combined treatment with the control treatment, the odds of reporting fake news increase significantly by 295 percent (i.e., it almost quadruples; medium effect; derived from the 0.253 odds ratio reported in Table 1 for the regression with the combined treatment as the base). At the same time, the odds of reporting fake news in the combined treatment is also significantly higher than in the injunctive SN treatment (101 percent increase in odds; small effect) and the descriptive SN treatment (150 percent increase in odds; medium effect). Thus, the data support H3.

In addition, the data reveal that the number of reports for real news is also significantly higher for the combined SN messages as compared to the absence of SN messages (control treatment; see Online Supplemental Appendix G). In the control treatment, three participants reported exactly one real news. In comparison, seven participants reported one real news in the combined treatment, and five participants reported two real news. For the descriptive SN only, we found an insignificant increase in real news reporting. It appears that adding the injunctive SN amplifies this effect.

As a robustness check, we conducted additional analyses to control for the participants' awareness of the reporting feature. Online Supplemental Appendix H provides detailed information.

H4: Effect of the Strength of Descriptive SN Messages on the Likelihood of Reporting News Stories as Fake

To evaluate the impact of the strength of a descriptive SN on the participants' reporting behavior, we use a logistic regression analysis. The binary dependent variable describes the reporting status of a post indicating whether the corresponding participant reported the post or not. The independent variable describes the strength of the descriptive SN. Since we cannot assume a direct connection between the numerical value of the strength of a descriptive SN and the influence on behavior, we interpret the descriptive SN's strength as an ordinal factor. We focus on the combined treatment since the effects of the descriptive SN are prominent. Since each of the 82 participants of this treatment has seen five fake news posts, our sample for this analysis consists of 410 fake news. We use the Huber-White procedure to account for clustering in observations that might arise from having multiple posts in a news feed [39, 93]. Table 2 summarizes the results. Online Supplemental Appendix I additionally varies the baseline.

The descriptive SN's strength increases the probability of fake news posts being reported (Table 2). We observe that, compared to no descriptive SN (strength 0), the odds are higher by 106 percent at a strength of 5, 128 percent at a strength of 25, 219 percent at a strength of 125, and 269 percent at a strength of 625. All these differences to the baseline are statistically significant; the effect sizes are small to medium. We observe indications that the weakest descriptive SN also exerts the least influence. For the strongest descriptive SN (strength 3,125),

Table 2. Results of the logistic regression analyzing the probability of a news post being reported based on the strength of the descriptive SNs.[a]

Strength	Coefficient	Standard Error	p-value		Odds Ratio (Effect Size)	Related Hypothesis
Intercept	-1.473	0.214	< 0.001	***	-	
5	0.723	0.313	0.021	*	2.061 (small)	
25	0.823	0.389	0.035	*	2.277 (small)	
125	1.160	0.329	< 0.001	***	3.190 (medium)	H4
625	1.306	0.311	< 0.001	***	3.691 (medium)	
3,125	0.570	0.353	0.106		1.768 (small)	
					Nagelkerke's R^2: 0.069	

Notes: [a]The baseline is a strength of 0. + $p < 0.1$; * $p < .05$; ** $p < .01$; *** $p < 0.001$. n = 410.

the odds are lower than for any other strength (see Table 2), statistically significantly lower than for the strength of 625 (Online Supplemental Appendix I), and not statistically significantly different from the odds for the baseline of no descriptive SN (Table 2). Thus, we observe that the probability of reporting a post increases to a strength of 625 and then drops for the strongest descriptive SN tested in our experiment. The ex-ante hypothesized explanation is that with the increasing number of reports from other users, one's own reporting action's perceived benefits reduced [94], leading to this inverted U-shape. Thus, the results support H4.

Nagelkerke's Pseudo R^2 for this logistic regression is low. Clearly, other factors beyond the descriptive SN affect reporting. Likely the content, the source, and other post-level characteristics influence reporting along with individual differences among participants. As each headline and news source is used only once, our data does not allow controlling for headline or source-effects. Also, we examined how the strength of a descriptive SN affects the reporting of real news. No significant differences could be observed (Online Supplemental Appendix I).

In summary, data from Study 1 supports hypotheses H1, H3, and H4. We do not find empirical support for H2 so far. A reason might be that for most participants, the specific implementation of the descriptive SN mixed elements of a positive and a negative descriptive SN. Thus, it is up to Study 2 to further test H2.

Study 2

Study 1 tested H2 without differentiation between positive and negative descriptive SNs. As mentioned above, seeing fake news posts flagged by other users can have a different effect than seeing real news posts flagged. Study 2 investigates this further.

Method

Study 2 is an online experiment similar to Study 1, with a notable difference in the treatments.

Participants

Recruiting procedures and filters regarding complete and valid data sets were like in Study 1, only using different Facebook groups to attract other participants (see Online Supplemental Appendix D for details). The sample comprises 157 participants who use social media more than once a week. Again, the sample is strongly composed of students (n = 116) with an average age of 25.6 years. About 60 percent of the participants are female (n = 94). In total, 142 participants state that they use social media as a news source at least once a week. Furthermore, 145 do not report any fake news or report a relatively minimal amount compared to their social environment. For more details, see Online Supplemental Appendix A.

Task

The task was the same as in Study 1. However, we modified the original fake news to decrease the level of authenticity further. Emoticons were added to two of the fake news, and the term "German Armed Forces Mission" changed to "War mission." We replaced one fake news completely because the participants in Study 1 did not recognize it well. We replaced the real news stories as they were outdated at the time of Study 2. These measures

further increased the difference between fake and real news so that the positive and negative descriptive SN is more readily experienced. See Online Supplemental Appendix D for details on the adjustments.

Treatments

We used three treatments: (1) *control treatment*, (2) *positive descriptive SN treatment* (where all five fake news but no real news was flagged with a descriptive SN message, with each strength occurring once), (3) *negative descriptive SN treatment* (where all five real news but no fake news was flagged with a descriptive SN message, with each strength occurring once).

Independent and Dependent Variables

Identical to Study 1.

Procedures

The procedures were mostly identical to Study 1 with two additions. First, in the fake assessment run, the participants assessed each news post (on a 5-step Likert scale) how fake or realistic it appears. Second, previous literature on social media users' motivation to report fake news is scarce. To gain further knowledge of it, we asked participants whether they report fake news if they see them. Based on their answer, participants were asked to rate a pre-selection of possible motivation factors or hurdles on a 5-level semantic differential and to add further factors in a free text. Online Supplemental Appendix E lists all survey scales used in Study 2.

Results

Studies 1 and 2 are comparable in terms of reporting behavior: Neither the number of reported fake news nor the number of reported real news is statistically significantly different between the control groups of both studies (Chi-square test, p-values of 0.511 and 0.272, respectively).

To analyze H2 in more detail, we compare—as in Study 1—participants' fake news reporting behavior in the different treatments by using an ordered logistic regression. Table 3 summarizes the results. We see that using a positive descriptive SN message leads to a 36 percent increase in odds compared to no SN message, but the effect size is less than small, and the effect is not statistically significant. Simultaneously, the results show that a negative descriptive SN message hardly leads to a change in fake news reporting behavior than no SN message since the odds increase by only 1 percent, and no significant effect is observed. Overall, even in the extreme setting of Study 2, no influence of the descriptive norm on the reporting behavior of fake news can be detected, which corroborates our results from Study 1. As in Study 1, we have conducted additional analysis to control for the participants' awareness of the reporting feature (see Online Supplemental Appendix H).

Additionally, Study 2 provides insights into motivation factors and hurdles of reporting fake news. Overall, 74 participants stated they would report fake news if they see them, and 83 would not. In summary, the motivation factors and hurdles to reporting are mainly intrinsically motivated and altruistic and not because of possible financial incentives. The answers from the free text have a considerable overlap with the preselected factors. However, two further hurdles could be identified. Table 4 shows the ranked results.

Table 3. Results of the ordered logistic regression to compare the three treatments with respect to fake news reporting in Study 2.[a]

Base Treatment			Coefficient	Standard Error	p-value		Odds Ratio (Effect Size)	Related Hypothesis
	Treatments	Positive descriptive SN	0.282	0.368	0.443		1.326 (less than small)	
		Negative descriptive SN	0.010	0.389	0.979		1.010 (less than small)	H2
Control	Intercepts	0 | 1	0.324	0.277	0.242			
		1 | 2	1.024	0.288	< 0.001	***		
		2 | 3	1.641	0.307	< 0.001	***		
		3 | 4	1.981	0.325	< 0.001	***		
		4 | 5	2.603	0.376	< 0.001	***		
							Nagelkerke's R^2: 0.005	

Notes: [a]The control group is the baseline. + $p < 0.1$; * $p < .05$; ** $p < .01$; *** $p < 0.001$. n = 157.

Table 4. Motivation factors (n = 74) and hurdles (n = 83) to report fake news measured on a 5-level scale where 1 corresponds to "strongly disagree" and 5 to "strongly agree."

Motivation factors and hurdles	Mean	Median	Standard Deviation
Preselected motivation factors to report fake news			
I would like to avoid negative consequences that result from the dissemination of fake news	4.662	5.0	0.556
It is important to me to help other people to form their opinion based on true facts	4.432	5.0	0.812
A correct news landscape is important to me	4.338	5.0	0.848
I would like to improve the living conditions for myself and other people	3.865	4.0	0.984
I do not want that the relevance of the topic will be reduced by the dissemination of fake news	3.797	4.0	1.085
Fake news causes an unpleasant feeling in me	3.757	4.0	1.083
I would like to improve the quality of the social media platform I use	3.527	4.0	1.230
I want to share my knowledge with others	2.973	3.0	1.085
I hope for appreciation from my social environment	1.811	1.5	0.989
I expect a material or financial incentive	1.716	1.0	1.040
Preselected hurdles to reporting fake news			
I do not believe that the act of reporting has any effect or counteracts the dissemination of fake news	3.807	4.0	1.109
I am only a consumer of news and do not actively participate in the public discourse	3.723	4.0	1.130
The procedure of reporting is too complex for me	3.157	3.0	1.366
Reporting fake news offers me no material or financial incentive	2.783	3.0	1.440
I am not interested in public opinion	2.687	3.0	1.278
Fake news serves for entertainment	2.350	2.0	1.204
I've never seen fake news	2.169	2.0	1.228
Fake news is subject to the freedom of opinion	2.120	2.0	1.005
I don't know how to report fake news	2.000	1.0	1.269
Further hurdles to report fake news (answers from free text question)			
Lack of responsibility			
Own uncertainty			

Discussion

Fake news is a greater threat than ever before, especially with the widespread and fast dissemination on social media. We aim to contribute to the emerging body of IS research on fake news by studying socio-technical interventions as remedies for fake news [8].

Specifically, we theoretically and experimentally analyze the effect of injunctive and descriptive SNs on social media users' reporting of fake news.

Our work leverages knowledge on fake news in social media and knowledge originating from social psychology on social norms and the focus theory of normative conduct. Our results suggest that SN messages impact the reporting behavior of users in online environments. Social media users can be guided in their behavior by highlighting desirable behavior and making transparent what other users are doing. However, caution is needed, as SNs can also give rise to the undesirable behavior of reporting more real news as "by-catch." Nevertheless, as the reporting levels of real news are substantially lower than the reporting levels of fake news, we argue that the overall effect of improving reporting by a combination of different SN messages is beneficial.

Social psychologists have already demonstrated the effectiveness of social norms. In part, our results are comparable to prior studies in different contexts regarding the effectiveness of injunctive SNs [7, 49, 60, 65, 70, 92]. Apart from that, our results are, in part, also contrary to the findings of other researchers that descriptive SNs are an appropriate tool to motivate desirable behavior by describing the behavior of others [16].

The combined application of both—injunctive and descriptive—SN messages has already led to previous studies' most substantial effects as it affects different motivations [9, 13, 14, 75]. We newly demonstrate that this is also the case in the domain of reporting fake news in social media and how SN messages can be integrated into social media interfaces. Following the focus theory of normative conduct [17], descriptive SN messages seem to focus users' attention on the injunctive SN (when this is sufficiently present) and thus improve users' reporting of fake news posts.

Also, the combined application of injunctive and descriptive SNs is more practical than just applying injunctive SNs. Our injunctive norm message has been displayed when the news feed was loaded and then clicked away by the user. This procedure is not very user-friendly; therefore, a social media provider would likely not show the message every time the platform is accessed. However, the descriptive SN message continually reminds the injunctive SN's user, thus redirecting the user's focus back to what is "ought to be done" [17].

Our results provide insights that indicate a threshold exists regarding the positive effects of a descriptive SN. As Wong et al. [94] already showed, the intention to report depends on the reporting action's perceived benefit. This finding does not fit with other studies that show a positive correlation of behavior and strength of a descriptive SN [21, 46]. However, these studies examined effects in scenarios such as online shopping [21], which promises a different individual benefit than reporting fake news on social media.

The general implication of our results for research is that researchers working on fake news should consider the effects of SNs. Researchers engaging in behavioral research should integrate SNs in their theorizing and examine the prevalent SNs in the populations they study to understand behavior better. Researchers engaging in design science research should be aware of the power of SNs to steer user behavior. They may leverage this power in their socio-technical designs and should be cautious of the side effects of invoking SNs unintentionally. Furthermore, to rigorously evaluate their designs, researchers should keep in mind that SNs can be an explanatory variable of user behavior in social media environments and, thus, might measure SNs to control for their effect. Finally, it is essential

to remember that publishing about phenomena like reporting behavior in social media also signals a descriptive SN and may affect user behavior.

Contribution to Theory

First, we provide theoretical and empirical support for SN messages' role as an effective approach to guiding social media users' reporting decisions toward reporting fake news, but SN messages are not always effective. On the one hand, injunctive SN messages are a motivational tool highlighting moral evaluation in social media environments and affecting user reporting behavior. On the other hand, empirically, we do not find a significant effect when we apply solely descriptive SN messages in social media environments. The combination of both types of SN messages is most effective.

Second, we provide further evidence that descriptive SNs can also backfire. Users in social media environments decrease in reporting fake news when a descriptive SN message's strength exceeds a threshold. The explanation may be that users lose their motivation to report fake news due to the reduced benefit, as many users have already reported the message.

Third, we find no empirical evidence that SN messages have a negative effect on the reporting of real news. Although our data do not show a significant effect, we cannot exclude that with increasing strength of a misplaced descriptive SN message, real news becomes more likely to be reported as fake news.

Recent IS research shows that social media interface design affects users' perception of and engagement with fake news [44, 45, 54, 56]. Moravec et al. showed that asking users to rate articles' truthfulness pushed them to think more critically about these articles' honesty and other articles [54]. In their study, the evaluation question was part of the experimental design and not intended as a potential permanent design element of a social media interface. Furthermore, flagging fake news as "disputed by 3rd party fact-checkers" [56, p. 1345] can increase users' cognitive activity – but the effectiveness depends on how the information is presented to the user [55, 56]. Furthermore, prior IS research suggests that providing different variants of source ratings positively affects users' cognitive and behavioral processes concerning fake news [44, 45, 54].

Our research on social norms to fight fake news is important as it is complementary to this prior IS research. We study how to make users evaluate the truthfulness with subtle changes to the social media interface while Moravec et al. examine this evaluation's effect once the user engages in it [54].

Furthermore, professional 3rd party fact-checking or even stronger measures like suppressing the occurrence of articles in users' news feeds or even deleting fake news from social media platforms complement the mechanisms of users reporting potential fakes to be flagged, suppressed, or deleted. While user reporting may be used for initially detecting fake news posts in the vast amount of social media content, subsequent reliance on expert knowledge to inspect suspicious posts may be important as truly judging a news story may require personal knowledge of the events in the news article.

The novel insight of our research is that social norm messages as part of the social media interface design can positively affect users' fake news reporting behavior. Our focus on users reporting potential fakes adds to the emergent body of IS research on socio-technical interventions that address the consumption of fake news [8]. This novel focus is essential

as it complements the portfolio of socio-technical interventions, which may lessen the challenges fake news poses for social media platform providers and society.

Practical Implications

Our results also have implications for practitioners who design social media platforms and interfaces. First, to encourage users to report fake news, messages should be displayed, highlighting the injunctive SN that this behavior is socially desired. Second, to further increase the number of reports for fake news, practitioners should consider combining injunctive and descriptive SN messages to focus their users' attention. In this, the use of descriptive SNs can also have adverse effects. The result is a trade-off between a higher number of reports for fake news and the accuracy of reporting fake or truthful news. One has to decide according to a social media platform's specific objectives, design, users, and environmental influences. For this reason, one may react from a certain threshold and choose alternative approaches. For example, starting from a defined number of reports, the exact number could be hidden, and an article or source rating could be shown.

Limitations and Future Research

Our findings are subject to limitations. The generalizability of our empirical results is limited by our design of the experimental materials (inspired by Facebook, a single wording for the injunctive SN message, a single way of displaying the descriptive SN message) and by our participants (young, German, educated, active social media users recruited via Facebook). Further, we cannot ensure that everyone has participated only once. However, no duplicates were identified among the 77 percent of participants in Study 1 and 76 percent of participants in Study 2 who provided an e-mail address for the voucher raffle. Additionally, in practice, social media users cannot always be asked to confirm the same injunctive SN message. A variation of the wording may help confront the user more often with injunctive SNs and promote desirable behavior. The motivational factors and hurdles for reporting fake news that we identified in Study 2 may support the development of injunctive SN messages. The composition of the posts' headlines, the perceived reputation of the posts' sources, the posts' images and subjects, and the users' familiarity with the news, might all affect the identification of fake news. We did not explore these factors related to detection but solely focused on reporting behavior.

In addition to these limitations, we also see five areas for future research emerging from our results. First, future research might empirically test our hypotheses with more diverse or complementary samples in other forms of social media that use other design elements or have a different structure, such as forums or instant messengers. In this context, future work should also deal with the boundary conditions of the descriptive SN in more detail. Our results suggest that there is an inflection point beyond which descriptive SNs no longer improve user behavior. Such inflection points should be examined in further empirical studies. Going beyond artificial experimental settings to field experiments on live social media platforms would greatly strengthen external validity. In doing so, researchers can develop a deeper understanding of the boundary conditions of the descriptive SN.

Second, future work should explore whether SNs from the immediate social environment have a more decisive influence, as prior research from IS [47] and other domains [7, 60] suggests. Furthermore, other factors interact with the cognition and judgment of users' news consumption in social media, for example, familiarity with news sources, topics, or headlines [45, 56].

Third, it appears a promising direction to study the effect of gamification on fake news reporting, that is "[…] the use of game design elements in non-game contexts" [23, p. 10] or "a process of enhancing a service with affordances for gameful experiences in order to support user's overall value creation" [41, p. 19]. Gamifying fake news reporting could take, for example, the form of social games [27] or badges [10] as investigated in other contexts by recent IS research.

Fourth, while SN-based interventions are socio-technical tools to foster active user engagement, which allows for follow-up measures such as user source rating or third-party fact-checking, we suspect that descriptive SNs also improve the ability to detect fake news. Future research may investigate this effect.

Fifth, another approach to improve users' behavior in digital environments is the concept of digital nudging [91]. IS researchers recently successfully modified the choice architecture to guide people, for example, through priming [22]. Hence, we assume that investigating the effectiveness of digital nudging approaches to fighting fake news in social media is promising.

Sixth, IS researchers should further develop descriptive and prescriptive knowledge about fake news in social media to fight this societal threat successfully. In addition to improving reporting behavior, it is vital to gain more insights into why users believe fake news, how can recognition be improved, and what countermeasures can be found to contain fake news? One possible direction could be to enhance the users' knowledge about specific topics by providing related but potentially contradicting articles below a social media post [33].

Conclusion

Overall, this paper investigates the influence of SNs on users reporting fake news on social media. We could theoretically derive and empirically validate that SNs can cause desirable behavioral changes by guiding users towards reporting fake news. The potential for applying these mechanisms seems considerable, as existing social media applications do not have to be adapted at great expense but can be extended by adding SN messages. Therefore, they are a promising tool to support existing methods against fake news and reduce their dissemination and negative impact.

References

1. Agresti, A.; and Kateri, M. Categorical data analysis. In M. Lovric (ed.), *International Encyclopedia of Statistical Science*. Berlin, Heidelberg: Springer, 2011.
2. Aisch, G; Huang, J; and Kang, C. Dissecting the #PizzaGate conspiracy theories, 2016. https://www.nytimes.com/interactive/2016/12/10/business/media/pizzagate.html   (accessed   on June 09, 2019).
3. Allcott, H.; and Gentzkow, M. *Social Media and Fake News in the 2016 Election*. Cambridge, MA: National Bureau of Economic Research, 2017.
4. Balmas, M. When fake news becomes real. *Communication Research*, *41*, 3 (2012), 430–454.

5. Barron, G. and Yechiam, E. Private e-mail requests and the diffusion of responsibility. *Computers in Human Behavior*, *18*, 5 (2002), 507–520.

6. Barthel, M; Mitchell, A; and Holcomb, J. Many Americans believe fake news is sowing confusion, 2016. http://www.journalism.org/2016/12/15/many-americans-believe-fake-news-is-sowing-confusion/(accessed on February 21 2020).

7. Baumgartner, S.E; Valkenburg, P.M; and Peter, J. The influence of descriptive and injunctive peer norms on adolescents' risky sexual online behavior. *Cyberpsychology, Behavior and Social Networking*, *14*, 12 (2011), 753–758.

8. Bernard, J.-G; Dennis, A.R; Galletta, D.F; Khan, A; and Webster, J. The tangled web: Studying online fake news. In H. Krcmar; J. Fedorowicz; W.F. Boh; J.M. Leimeister; and S. Wattal (eds.), *Proceedings of the 40th International Conference on Information Systems, ICIS 2019*: Association for Information Systems, 2019.

9. Bernedo, M; Ferraro, P.J; and Price, M. The persistent impacts of norm-based messaging and their implications for water conservation. *Journal of Consumer Policy*, *37*, 3 (2014), 437–452.

10. Bhattacharyya, S; Banerjee, S; Bose, I; and Kankanhalli, A. Temporal effects of repeated recognition and lack of recognition on online community contributions. *Journal of Management Information Systems*, *37*, 2 (2020), 536–562.

11. Blair, C.A; Foster Thompson, L; and Wuensch, K.L. Electronic helping behavior: The virtual presence of others makes a difference. *Basic and Applied Social Psychology*, *27*, 2 (2005), 171–178.

12. Chatfield, A.T; Reddick, C.G; and Choi, K.P. Online media use of false news to frame the 2016 Trump Presidential Campaign. In C.C. Hinant and A. Ojo (eds.), *Proceedings of the 18th Annual International Conference on Digital Government Research*, 2017, pp. 213–222.

13. Cialdini, R.B. Crafting normative messages to protect the environment. *Current Directions in Psychological Science*, *12*, 4 (2003), 105–109.

14. Cialdini, R.B. Descriptive social norms as underappreciated sources of social control. *Psychometrika*, *72*, 2 (2007), 263–268.

15. Cialdini, R.B; Demaine, L.J; Sagarin, B.J; Barrett, D.W; Rhoads, K; and Winter, P.L. Managing social norms for persuasive impact. *Social Influence*, *1*, 1 (2006), 3–15.

16. Cialdini, R.B; Kallgren, C.A; and Reno, R.R. A focus theory of normative conduct: A theoretical refinement and reevaluation of the role of norms in human behavior. In M.P. Zanna (ed.), *Advances in Experimental Social Psychology Volume 24*: Elsevier, 1991, pp. 201–234.

17. Cialdini, R.B; Reno, R.R; and Kallgren, C.A. A focus theory of normative conduct: Recycling the concept of norms to reduce littering in public places. *Journal of Personality and Social Psychology*, *58*, 6 (1990), 1015–1026.

18. Clarke, J; Chen, H; Du, D; and Hu, Y.J. Fake news, investor attention, and market reaction. *Information Systems Research*, (2020), online ahead of print.

19. Cohen, J. *Statistical Power Analysis for the Behavioral Sciences*. New York, NY: Academic Press, 1969.

20. Croson, R; Handy, F; and Shang, J. Keeping up with the Joneses: The relationship of perceived descriptive social norms, social information, and charitable giving. *Nonprofit Management Leadership*, *19*, 4 (2009), 467–489.

21. Demarque, C; Charalambides, L; Hilton, D.J; and Waroquier, L. Nudging sustainable consumption: The use of descriptive norms to promote a minority behavior in a realistic online shopping environment. *Journal of Environmental Psychology, 43* (2015), 166–174.

22. Dennis, A.R; Yuan, L; Feng, X; Webb, E; and Hsieh, C.J. Digital nudging: Numeric and semantic priming in e-commerce. *Journal of Management Information Systems*, *37*, 1 (2020), 39–65.

23. Deterding, S; Dixon, D; Khaled, R; and Nacke, L. From game design elements to gamefulness: Defining gamification. In A. Lugmayr; H. Franssila; C. Safran; and I. Hammouda (eds.), *Proceedings of the 15th International Academic MindTrek Conference Envisioning Future Media Environments*. New York, NY: ACM, 2011, pp. 9–15.

24. Deutsch, M.; and Gerard, H.B. A study of normative and informational social influences upon individual judgment. *The Journal of Abnormal and Social Psychology*, *51*, 3 (1955), 629–636.

25. DiFranzo, D; Taylor, S.H; Kazerooni, F; Wherry, O.D; and Bazarova, N.N. Upstanding by design. In R. Mandryk; M. Hancock; M. Perry; and A. Cox (eds.), *Proceedings of the 2018 CHI Conference on Human Factors in Computing* Systems - *CHI '18*. New York, New York, USA: ACM Press, 2018, pp. 1–12.

26. Facebook. Company Info, 2019. https://newsroom.fb.com/company-info/ (accessed on January 03, 2019).

27. Fang, B; Zheng, Z; Ye, Q; and Goes, P.B. Social Influence and Monetization of Freemium Social Games. *Journal of Management Information Systems*, *36*, 3 (2019), 730–754.

28. Ferrara, E; Varol, O; Davis, C; Menczer, F; and Flammini, A. The rise of social bots. *Communications of the ACM*, *59*, 7 (2016), 96–104.

29. Festinger, L. A theory of cognitive dissonance. Stanford: CA: Stanford University Press, 1957.

30. Fischer, P; Krueger, J.I; Greitemeyer, T; Vogrincic, C; Kastenmüller, A; Frey, D; Heene, M; Wicher, M; and Kainbacher, M. The bystander-effect: A meta-analytic review on bystander intervention in dangerous and non-dangerous emergencies. *Psychological Bulletin*, *137*, 4 (2011), 517–537.

31. Flynn, D.J; Nyhan, B; and Reifler, J. The nature and origins of misperceptions: Understanding false and unsupported beliefs about politics. *Advances in Political Psychology*, *38*, 2 (2017), 127–150.

32. Gerber, A.S.; and Rogers, T. Descriptive social norms and motivation to vote: Everybody's voting and so should you. *The Journal of Politics*, *71*, 1 (2009), 178–191.

33. Gimpel, H; Heger, S; Kasper, J; and Schäfer, R. The power of related articles – Improving fake news detection on social media platforms. In T. Bui (ed.), *Proceedings of the 53rd Hawaii International Conference on System Sciences*: Hawaii International Conference on System Sciences, 2020.

34. Goldstein, N.J; Cialdini, R.B; and Griskevicius, V. A room with a viewpoint: Using social norms to motivate environmental conservation in hotels. *Journal of Consumer Research*, *35*, 3 (2008), 472–482.

35. Gottfried, J.; and Shearer, E. News use across social media platforms, 2016. https://assets. pewresearch.org/wp-content/uploads/sites/13/2016/05/PJ_2016.05.26_social-media-and-news_FINAL-1.pdf (accessed on February 21, 2020).

36. Greenhill, K.M.; and Oppenheim, B. Rumor has it: The adoption of unverified information in conflict zones. *International Studies Quarterly*, *61*, 3 (2017), 660–676.

37. Hasher, L; Goldstein, D; and Toppino, T. Frequency and the conference of referential validity. *Journal of Verbal Learning and Verbal Behavior*, *16*, 1 (1977), 107–112.

38. Housholder, E.E.; and LaMarre, H.L. Facebook politics: Toward a process model for achieving political source credibility through social media. *Journal of Information Technology & Politics*, *11*, 4 (2014), 368–382.

39. Huber, P.J. The behavior of maximum likelihood estimates under nonstandard conditions. In L.M. Le Cam and J. Neyman (eds.), *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*. Berkeley, CA: University of California Press, 1967, pp. 221–233.

40. Hunt, E. "Disputed by Multiple Fact-Checkers": Facebook Rolls Out New Alert to Combat Fake News, 2017. https://www.theguardian.com/technology/2017/mar/22/facebook-fact-checking-tool-fake-news (accessed on January 22, 2019).

41. Huotari, K.; and Hamari, J. Defining gamification - A service marketing perspective. In A. Lugmayr (ed.), *Proceedings of the 16th International Academic MindTrek Conference*. New York, NY: ACM, 2012, pp. 17–22.

42. Jacobson, R.P; Mortensen, C.R; and Cialdini, R.B. Bodies obliged and unbound: Differentiated response tendencies for injunctive and descriptive social norms. *Journal of Personality and Social Psychology*, *100*, 3 (2011), 433–448.

43. Kaplan, A.M.; and Haenlein, M. Users of the world, unite!: The challenges and opportunities of social media. *Business Horizons*, *53*, 1 (2010), 59–68.

44. Kim, A.; and Dennis, A.R. Says who? The effects of presentation format and source rating on fake news in social media. *MIS Quarterly*, *43*, 3 (2019), 1025–1039.

45. Kim, A; Moravec, P.L; and Dennis, A.R. Combating fake news on social media with source ratings: The effects of user and expert reputation ratings. *Journal of Management Information Systems*, *36*, 3 (2019), 931–968.

46. Kormos, C; Gifford, R; and Brown, E. The influence of descriptive social norm information on sustainable transportation behavior. *Environment and Behavior*, *47*, 5 (2014), 479–501.

47. Kuem, J; Khansa, L; and Kim, S.S. Prominence and engagement: Different mechanisms regulating continuance and contribution in online communities. *Journal of Management Information Systems*, *37*, 1 (2020), 162–190.

48. Laato, S; Islam, A.K.M.N; Islam, M.N; and Whelan, E. What drives unverified information sharing and cyberchondria during the COVID-19 pandemic? *European Journal of Information Systems*, *29*, 3 (2020), 288–305.

49. LaBrie, J.W; Hummer, J.F; Neighbors, C; and Larimer, M.E. Whose opinion matters? The Relationship between injunctive norms and alcohol consequences in college students. *Addictive Behaviors*, *35*, 4 (2010), 343–349.

50. Latané, B.; and Darley, J.M. *The Unresponsive Bystander: Why Doesn't He Help?* New York: Appleton-Century Crofts, 1970.

51. Lazer, D.M.J; Baum, M.A; Benkler, Y; Berinsky, A.J; Greenhill, K.M; Menczer, F; Metzger, M. J; Nyhan, B; Pennycook, G; Rothschild, D; Schudson, M; Sloman, S.A; Sunstein, C.R; Thorson, E.A; Watts, D.J; and Zittrain, J.L. The science of fake news. *Science*, *359*, 6380 (2018), 1094–1096.

52. Lee, C.F; Lee, J.C; and Lee, A.C. *Statistics for Business and Financial Economics*, 2nd edn. Singapore: World Scientific, 2000.

53. Mills, J. Improving The 1957 version of dissonance theory. In E. Harmon-Jones (ed.), *Cognitive Dissonance: Reexamining a Pivotal Theory in Psychology*, 2nd edn. Washington: American Psychological Association, 2019, pp. 27–39.

54. Moravec, P; Kim, A; Dennis, A; and Minas, R. Do you really know if it's true? How asking users to rate stories affects belief in fake news on social media. In T. Bui (ed.), *Proceedings of the 52nd Hawaii International Conference on System Sciences*: Hawaii International Conference on System Sciences, 2019.

55. Moravec, P.L; Kim, A; and Dennis, A.R. Appealing to sense and sensibility: System 1 and system 2 interventions for fake news on social media. *Information Systems Research, 31*, 3 (2020), 987-1006.

56. Moravec, P.L; Minas, R; and Dennis, A.R. Fake news on social media: people believe what they want to believe when it makes no sense at all. *MIS Quarterly*, *43*, 4 (2019), 1343–1360.

57. Mosseri, A. News Feed FYI: Addressing Hoaxes and Fake News, 2016. https://newsroom.fb. com/news/2016/12/news-feed-fyi-addressing-hoaxes-and-fake-news/(accessed on January 22, 2019).

58. Murray, A. How to Report Fake News to Social Media, 2016. https://www.bbc.com/news/ 38053324 (accessed on January 04, 2019).

59. Mustafaraj, E.; and Metaxas, P.T. The fake news spreading plague. In P. Boldi; K. Kinder-Kurlanda; P. Fox; D. McGuinness; and L. Poirer (eds.), *Proceedings of the 2017 ACM on Web Science Conference - WebSci '17*. New York, USA: ACM Press, 2017, pp. 235–239.

60. Neighbors, C; Lostutter, T.W; Whiteside, U; Fossos, N; Walker, D.D; and Larimer, M.E. Injunctive norms and problem gambling among college students. *Journal of Gambling Studies*, *23*, 3 (2007), 259–273.

61. Newman, N; Fletcher, R; Kalogeropoulos, A; Levy, D.A.L; and Nielsen, R.K. Reuters Institute Digital News Report 2017, 2017. https://reutersinstitute.politics.ox.ac.uk/sites/default/files/ Digital%20News%20Report%202017%20web_0.pdf (accessed on October 27, 2020).

62. Nigbur, D; Lyons, E; and Uzzell, D. Attitudes, norms, identity and environmental behaviour: Using an expanded theory of planned behaviour to predict participation in a kerbside recycling programme. *The British Journal of Social Psychology*, *49*, 2 (2010), 259–284.

63. Nyhan, B.; and Reifler, J. When corrections fail: The persistence of political misperceptions. *Political Behavior*, *32*, 2 (2010), 303–330.

64. Oxford Internet Institute, University of Oxford. Resource For Understanding Political Bots, 2016. https://www.oii.ox.ac.uk/blog/resource-for-understanding-political-bots/(accessed on February 21, 2020).

65. Park, H.S.; and Smith, S.W. Distinctiveness and influence of subjective norms, personal descriptive and injunctive norms, and societal descriptive and injunctive norms on behavioral intent: A case of two behaviors critical to organ donation. *Human Communication Research*, *33*, 2 (2007), 194–218.

66. Pempek, T.A; Yermolayeva, Y.A; and Calvert, S.L. College students' social networking experiences on Facebook. *Journal of Applied Developmental Psychology*, *30*, 3 (2009), 227–238.

67. Pennycook, G; Cannon, T.D; and Rand, D.G. Prior exposure increases perceived accuracy of fake news. *Journal of Experimental Psychology. General*, *147*, 12 (2018), 1865–1880.

68. Petty, R.E.; and Cacioppo, J.T. Elaboration likelihood model. In L. Berkowitz (ed.), *Advances in Experimental Social Psychology*. San Diego, USA: Academic Press, 1986, pp. 123–205.

69. Reno, R.R; Cialdini, R.B; and Kallgren, C.A. The transsituational influence of social norms. *Journal of Personality and Social Psychology*, *64*, 1 (1993), 104–112.

70. Rimal, R.N.; and Real, K. Understanding the influence of perceived norms on behaviors. *Communication Theory*, *13*, 2 (2003), 184–203.

71. Robinson, E; Fleming, A; and Higgs, S. Prompting healthier eating: Testing the use of health and social norm based messages. *Health Psychology: Official Journal of the Division of Health Psychology, American Psychological Association*, *33*, 9 (2014), 1057–1064.

72. Rubin, V.L; Chen, Y; and Conroy, N.J. Deception detection for news: Three types of fakes. In *Proceedings of the Association for Information Science and Technology*, 2016, pp. 1–4.

73. Sánchez-Meca, J; Marín-Martínez, F; and Chacón-Moscoso, S. Effect-size indices for dichotomized outcomes in meta-analysis. *Psychological Methods*, *8*, 4 (2003), 448–467.

74. Schmehl, K; and Lytvynenko, J. 7 Out of the 10 most viral articles about Angela Merkel On Facebook Are False, 2017. https://www.buzzfeednews.com/article/karstenschmehl/top-merkel-news (accessed on October 07, 2020).

75. Schultz, P.W; Khazian, A.M; and Zaleski, A.C. Using normative social influence to promote conservation among hotel guests. *Social Influence*, *3*, 1 (2008), 4–23.

76. Schultz, P.W; Nolan, J.M; Cialdini, R.B; Goldstein, N.J; and Griskevicius, V. The constructive, destructive, and reconstructive power of social norms. *Psychological Science*, *18*, 5 (2007), 429–434.

77. Schwarz, N; Sanna, L.J; Skurnik, I; and Yoon, C. Metacognitive experiences and the intricacies of setting people straight: implications for debiasing and public information campaigns. In M. P. Zanna (ed.), *Advances in Experimental Social Psychology*, 1st edn.: Elsevier Textbooks, 2007, pp. 127–161.

78. Shao, C; Ciampaglia, G.L; Flammini, A; and Menczer, F. Hoaxy: A Platform for Tracking Online Misinformation. In J. Bourdeau; J.A. Hendler; R.N. Nkambou; I. Horrocks; and B.Y. Zhao (eds.). *WWW'16 Companion: Proceedings of the 25th International Conference on World Wide Web May 11-15, 2016, Montreal, Canada*. Republic and Canton of Geneva: International World Wide Web Conferences Steering Committee, 2016, pp. 745–750.

79. Shu, K; Sliva, A; Wang, S; Tang, J; and Liu, H. Fake news detection on social media. *ACM SIGKDD Explorations Newsletter*, *19*, 1 (2017), 22–36.

80. Shukla, S.; and Lyons, T. Blocking ads from pages that repeatedly share false news, 2017. https://newsroom.fb.com/news/2017/08/blocking-ads-from-pages-that-repeatedly-share-false-news/(accessed on January 22, 2019).

81. Silverman, C. Here are 50 of the biggest fake news hits on Facebook from 2016, 2016. https://www.buzzfeed.com/craigsilverman/top-fake-news-of-2016 (accessed on January 22, 2019).

82. Simon, H.A. Rational decision making in business organizations. *The American Economic Review*, *69*, 4 (1979), 493–513.

83. The Editorial Board. Facebook and the Digital Virus Called Fake News, 2016. https://www.nytimes.com/2016/11/20/opinion/sunday/facebook-and-the-digital-virus-called-fake-news.html (accessed on February 21, 2020).

84. Twitter International Company. Q3 2018 Letter to Shareholders, 2018. https://s22.q4cdn.com/826641620/files/doc_financials/2018/q3/Q3-2018-Shareholder-Letter.pdf (accessed on January 03, 2019).

85. van der Linden, S; Leiserowitz, A; Rosenthal, S; and Maibach, E. Inoculating the public against misinformation about climate change. *Global Challenges*, *1*, 2 (2017), 1–7.

86. Voelpel, S.C; Eckhoff, R.A; and Förster, J. David against Goliath?: Group size and bystander effects in virtual knowledge sharing. *Human Relations*, *61*, 2 (2008), 271–295.

87. Vosoughi, S; Roy, D; and Aral, S. The spread of true and false news online. *Science*, *359*, 6380 (2018), 1146–1151.

88. Warzel, C.; and Mac, R. Twitter just launched a midterm elections page and it's already full of garbage, 2018. https://www.buzzfeednews.com/article/charliewarzel/twitter-just-launched-a-midterms-page-and-its-already (accessed on January 03, 2019).

89. Wathen, C.N.; and Burkell, J. Believe it or not: Factors influencing credibility on the Web. *Journal of the American Society for Information Science and Technology*, *53*, 2 (2002), 134–144.

90. Weedon, J; Nuland, W; and Stamos, A. Information Operations and Facebook, 2017. https://fbnewsroomus.files.wordpress.com/2017/04/facebook-and-information-operations-v1.pdf (accessed on January 22, 2019).

91. Weinmann, M; Schneider, C; and Vom Brocke, J. Digital nudging. *Business & Information Systems Engineering*, *58*, 6 (2016), 433–436.

92. Wenzel, M. The social side of sanctions: Personal and social norms as moderators of deterrence. *Law and Human Behavior*, *28*, 5 (2004), 547–567.

93. White, H. A Heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica: Journal of the Econometric Society* (1980), 817–838.

94. Wong, R.Y.M; Cheung, C.M.K; and Xiao, B. Combating online abuse: What drives people to use online reporting functions on social networking sites. In T. Bui and R.H. Sprague (eds.), *Proceedings of the 49th Hawaii International Conference on System Sciences*, 2016, pp. 415–424.

About the Authors

Henner Gimpel (henner.gimpel@fim-rc.de; corresponding author) holds the chair for Digital Management at the University of Hohenheim, Germany. He holds a doctorate in Information Systems from the Karlsruhe Institute of Technology. Dr. Gimpel is a member of the Research Center Finance & Information Management and the Project Group Business & Information Systems Engineering of Fraunhofer FIT, Germany. The main focus of his work is the analysis and design of the use of digital technologies and media in the work and private lives of individuals and groups.

Sebastian Heger (sebastian.heger@fim-rc.de) holds a doctorate in Information Systems from the University of Augsburg. His research focuses on the topics of socio-technical information systems and sustainability.

Christian Olenberger (christian.olenberger@fim-rc.de) received his M. S. with honors in the elite graduate program Finance & Information Management at the Technical University of Munich, the University of Augsburg, and the University of Bayreuth, Germany. He is a doctoral candidate, focusing primarily on the digitization and information systems design.

Lena Utz (lena.utz@fim-rc.de) received her M. S. with honors in the elite graduate program Finance & Information Management at the Technical University of Munich, the University of Augsburg, and the University of Bayreuth, Germany. She is a doctoral candidate at the FIM Research Center, Germany, focusing on customer relationship management and human information behavior.